

## CLAIMS

What is claimed is:

1. A system that facilitates mitigation of outgoing spam, comprising:  
a detection component that detects a potential spammer in connection with at least one outgoing message, the outgoing message comprising at least one of instant message spam, whisper spam, and chat room spam, the detection being based in part on at least one of spam filtering, message volume monitoring, total recipient counting, unique recipient counting, message rate monitoring, number of apparently legitimate messages, and number of non-deliverable messages; and  
an action component that upon receiving information from the detection component that an entity is a potential spammer, initiates at least one action that facilitates any one of confirming that the entity is a spammer, mitigating spamming by the entity, increasing spammer cost, and a combination thereof.
2. The system of claim 1, the outgoing message further comprising email message spam.
3. The system of claim 1 wherein the action initiated comprises at least one of:  
shutting down the potential spammer's user account;  
requiring any one of a HIP challenge and a computational challenge to be solved by the potential spammer and the potential spammer computer, respectively;  
sending the potential spammer a legal notice regarding at least one violation of messaging service terms; and  
manual inspection of at least a subset of outgoing messages generated by the potential spammer.
4. The system of claim 1, wherein message volume monitoring comprises at least one of tracking and counting outgoing messages.

5. The system of claim 1, wherein the recipient count is computed with each recipient counted only once.

6. The system of claim 5, comprising keeping track of the maximum score per recipient.

7. The system of claim 5, comprising using a pseudo-random function of recipients to estimate the recipient count, or related scores.

8. The system of claim 1, wherein the message rate monitoring comprises computing the volume of outgoing messages over a duration of time.

9. The system of claim 8, wherein the duration of time comprises at least one of minutes, hours, days, weeks, months, and years.

10. The system of claim 1, wherein the message volume monitoring comprises a total volume of messages since activation of a user account.

11. The system of claim 1, wherein each recipient of an outgoing message constitutes one message.

12. The system of claim 1, wherein the recipient count comprises one or more recipients listed in at least one of a to: field, a cc: field, and a bcc: field.

13. The system of claim 1, wherein the detection component processes and analyzes the outgoing messages to determine at least one of whether the message is likely to be spam and whether the sender is a potential spammer.

14. The system of claim 1, wherein the number of apparently legitimate messages is used as a bonus to offset other scores.

15. The system of claim 14, wherein the number of apparently legitimate messages is estimated with a spam filter.

16. The system of claim 14, wherein the bonus from the number of apparently legitimate messages is limited

17. The system of claim 1, wherein the number of non-deliverable messages is estimated at least in part from failures at message delivery time

18. The system of claim 1, wherein the number of non-deliverable messages is estimated at least in part from Non Delivery Receipts.

19. The system of claim 18, wherein validity of the Non Delivery Receipts is checked.

20. The system of claim 19, wherein validity of the Non Delivery Receipts is checked against a list of recipients of messages from the sender.

21. The system of claim 20, wherein the list of recipients is a sample and the penalty of a Non Delivery Receipt is correspondingly increased.

22. The system of claim 1, wherein the detection component computes scores assigned to the outgoing messages to determine a total score per sender and compares the total score per sender with at least one threshold level to ascertain whether the sender is a potential spammer.

23. The system of claim 22, wherein threshold levels are adjustable per sender.

24. The system of claim 1, wherein spam filtering comprises employing a filter trained to recognize at least one of non-spam like features and spam-like features in outgoing messages.

25. The system of claim 1, wherein spam filtering is performed with a machine learning approach.

26. The system of claim 1, wherein spam filtering comprises assigning a probability per outgoing message to indicate a likelihood that the message is any one of more spam-like or less spam-like.

27. The system of claim 1, further comprising a scoring component that operates in connection with at least one of the spam filtering, total recipient count, unique recipient count, message volume monitoring and message rate monitoring.

28. The system of claim 27, wherein the scoring component assigns a score per sender based at least in part upon volume of outgoing messages, rate of outgoing messages, recipient count, and message content.

29. The system of claim 27, wherein the scoring component assigns and/or adds a constant value to one or more outgoing messages to mitigate spammers from manipulating spam filtering systems.

30. The system of claim 27, wherein the scoring component assigns a selected value to outgoing messages identified as having at least one spam-like feature.

31. The system of claim 30, wherein the at least one spam-like feature is a URL.

32. The system of claim 30, wherein the at least one spam-like feature comprises contact information.

33. The system of claim 32, wherein the contact information comprises a telephone number, the telephone number comprising at least one of an area code and a prefix to identify a geographic location associated with the message to thereby facilitate identifying the potential spammer.

34. The system of claim 1, further comprising a user-based message generator component that generates outgoing messages addressed to one or more recipients based in part upon sender preferences.

35. A method that facilitates mitigation of outgoing spam comprising:  
detecting a potential spammer in connection with at least one outgoing message, the outgoing message comprising at least one of instant message spam, whisper spam, and chat room spam, the detection being based in part on at least one of spam filtering, message volume monitoring, total recipient counting, unique recipient counting, and message rate monitoring;

receiving information from the detection component that an entity is a potential spammer; and

initiating at least one action that facilitates any one of confirming that the entity is a spammer, mitigating spamming by the entity, and increasing spammer cost.

36. The method of claim 35, wherein the at least one outgoing message further comprises mail message spam.

37. The method of claim 35, further comprising monitoring outgoing messages per sender with respect to at least one of a volume of outgoing messages, a volume of recipients, and a rate of outgoing messages.

38. The method of claim 35, wherein detecting a potential spammer comprises:
- performing at least one of the following:
    - assigning a score per outgoing message based at least in part upon content of the message;
    - assigning a score per sender based at least in part upon outgoing message volume per sender;
    - assigning a score per sender based at least in part upon outgoing message rate per sender;
    - assigning a score per sender based at least in part upon a total recipient count per sender; and
    - assigning a score per sender based at least in part upon a unique recipient count per sender;
  - computing a total score per sender; and
  - determining whether the sender is a potential spammer based at least in part upon the total score associated with the sender.
39. The method of claim 38, wherein the total score exceeds a threshold level which thereby indicates that the respective sender is at least a potential spammer.
40. The method of claim 35, further comprising tracking one or more recipients and associated outgoing messages addressed to the recipients to facilitate identifying one or more most spam-like messages received per sender.
41. The method of claim 40, further comprising assigning one or more scores to the one or more most spam-like messages and aggregating the scores per sender to compute a total score per sender.
42. The method of claim 35, wherein the at least one action comprises terminating the sender account.

43. The method of claim 42, wherein the sender account is terminated when there is substantial certainty that the outgoing messages sent by a sender are spam.

44. The method of claim 43, wherein substantial certainty that the outgoing messages are spam is determined in part by at least one of the following:

at least a portion of the outgoing message comprises at least one of an exact match and a near match to known spam;

at least a portion of the outgoing message comprises a phrase that a human has determined to be spam-like;

a probability assigned by a spam filtering filter exceeds at least one threshold level; and

a message sent for human inspection is determined to be spam.

45. The method of claim 35, wherein the at least one action comprises temporarily suspending outgoing message delivery from the sender account.

46. The method of claim 35, wherein the at least one action comprises requiring the sender account to resolve one or more challenges.

47. The system of claim 44, wherein the account is volume limited per challenge until a determined number of challenges are solved, and is then rate limited thereafter.

48. The system of claim 45, wherein the rate limit may be increased by solving additional challenges.

49. The method of claim 46, wherein the one or more challenges comprise a computational challenge or a human interactive proof.

50. The method of claim 46, wherein the one or more challenges are delivered as a pop up message.

51. The method of claim 46, wherein the one or more challenges are delivered to the sender account *via* a message format similar to the sender's outgoing messages.

52. The method of claim 46, wherein the one or more challenges are delivered to the sender account in response to feedback from a server that a shutdown of the account is approaching.

53. The method of claim 35, wherein the at least one action comprises sending a legal notice that the sender is in violation of terms of service.

54. The method of claim 53, comprising responding to the legal notice *via* at least one of providing an electronic signature and clicking on a link.

55. The method of claim 53, wherein the legal notice is delivered *via* a pop-up message.

56. The method of claim 35, wherein delivery of outgoing messages is temporarily suspended until a response to the action is received.

57. The method of claim 35, wherein a minimum number of outgoing messages are permitted for delivery before a response to the action is received.

58. The method of claim 35, further comprising estimating a total volume of recipients per sender to facilitate identifying a potential spammer.

59. The method of claim 58, wherein estimating a total volume of distinct recipients per sender comprises:

computing a hash function per recipient to obtain a hash value per recipient;

setting a hash modulo value; and

adding the recipient to a list for message tracking when the recipient's hash value equals the hash modulo value to facilitate estimating a total volume of distinct recipients per sender.

60. The method of claim 59, further comprising:  
tracking worst-scoring messages each listed recipient receives per sender;  
computing a total score of substantially all listed recipients' scores per sender; and  
comparing the total score per sender with a threshold level associated with the sender to determine whether the sender is a potential spammer.

61. A method that facilitates periodic validation of non-spammer like activity by a user account comprising:  
monitoring at least one of a volume of outgoing messages, a volume of recipients, a rate of outgoing messages;  
requiring the user account to resolve one or more challenges after at least one of a number of outgoing messages are counted and a number of recipients are counted; and  
suspending delivery of subsequent outgoing messages until the one or more challenges are resolved.

62. The method of claim 61, wherein each recipient listed in a message counts as an individual message.

63. The method of claim 61, wherein the challenge is a computational challenge.

64. The method of claim 61, wherein the challenge is a human interactive proof.

65. A method of mitigating spam comprising:

performing at least one economic analysis to determine sender volume limits based at least in part on spammer behavior and legitimate user behavior; and

limiting the sender volume to at least one of:

a maximum number per challenge resolved; and

a maximum number per fee paid by a sender.

66. The method of claim 65, wherein the challenge is at least one of a human interactive proof and a computational challenge.

67. The method of claim 65, wherein the fee is any one of a user account set up fee, a monthly fee, a per-outgoing message fee, and a per number of outgoing messages fee.

68. The method of claim 65, wherein the fee is limited to an amount that is low enough for legitimate users to willingly pay and high enough to mitigate sending spam messages.

69. The method of claim 65, wherein a sender volume limit restricts a number of outgoing messages over a duration of time.

70. A computer readable medium comprising the method of claim 1.

71. A computer-readable medium having stored thereon the following computer executable components:

a detection component that detects a potential spammer in connection with at least one outgoing message, the outgoing message comprising at least one of instant message spam, whisper spam, and chat room spam, the detection being based in part on at least one of spam filtering, message volume monitoring, recipient counting, and message rate monitoring; and

an action component that upon receiving information from the detection component that an entity is a potential spammer, initiates at least one action that

facilitates any one of confirming that the entity is a spammer, mitigating spamming by the entity, increasing spammer cost, and a combination thereof.

72. A data packet adapted to be transmitted between two or more computer processes facilitating identify potential spammers, the data packet comprising:

information associated with detecting spam-like characteristics with at least one outgoing message, the outgoing message comprising at least one of instant message spam, whisper spam, and chat room spam, the detection being based in part on at least one of spam filtering, message volume monitoring, recipient counting, and message rate monitoring, wherein the information determines whether to initiate at least one action that facilitates any one of confirming that the entity is a spammer, mitigating spamming by the entity, and increasing spammer cost.

73. A system that facilitates spam detection comprising:

a means for detecting a potential spammer in connection with at least one outgoing message, the outgoing message comprising at least one of instant message spam, whisper spam, and chat room spam, the detection being based in part on at least one of spam filtering, message volume monitoring, recipient counting, and message rate monitoring;

a means for receiving information from the detection component that an entity is a potential spammer; and

a means for initiating at least one action that facilitates any one of confirming that the entity is a spammer, mitigating spamming by the entity, and increasing spammer cost.

74. A system that facilitates periodic validation of non-spammer like activity by a user account comprising:

a means for monitoring at least one of a volume of outgoing messages, a volume of recipients, a rate of outgoing messages;

a means for requiring the user account to resolve one or more challenges after at least one of a number of outgoing messages are counted and a number of recipients are counted; and

a means for suspending delivery of subsequent outgoing messages until the one or more challenges are resolved.

75. A system that facilitates mitigating spam comprising:

a means for performing at least one economic analysis to determine sender volume limits based at least in part on spammer behavior and legitimate user behavior; and

a means for limiting the sender volume to at least one of:

a maximum number per challenge resolved; and

a maximum number per fee paid by a sender.